Interobserver variation in CD30 immunohistochemistry interpretation; consequences for patient selection for targeted treatment

Koens, Lianne; van de Ven, Peter M.; Hijmering, Nathalie J.; Kersten, Marie J.; Diepstra, Arjan; Chamuleau, Martine; de Jong, Daphne

Published in:
Histopathology

DOI:
10.1111/his.13647

Histopathology

Histopathology 2018, **73**, 473–482. DOI: 10.1111/his.13647

Interobserver variation in CD30 immunohistochemistry interpretation; consequences for patient selection for targeted treatment

Lianne Koens,[1] Peter M van de Ven,[2] Nathalie J Hijmering,[3] Marie J Kersten,[4] Arjan Diepstra,[5] Martine Chamuleau[6] & Daphne de Jong[3]

[1]*Department of Pathology, Academic Medical Center,* [2]*Department of Epidemiology and Biostatistics,* [3]*Department of Pathology, VU Medical Center,* [4]*Department of Hematology, Academic Medical Center, Amsterdam,* [5]*Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, and* [6]*Department of Hematology, VU Medical Center, Amsterdam, the Netherlands*

Date of submission 20 February 2018
Accepted for publication 8 May 2018
Published online *Article Accepted* 14 May 2018

Koens L, van de Ven P M, Hijmering N J, Kersten M J, Diepstra A, Chamuleau M & de Jong D (2018) *Histopathology* **73**, 473–482. https://doi.org/10.1111/his.13647

Interobserver variation in CD30 immunohistochemistry interpretation; consequences for patient selection for targeted treatment

Aims: CD30 immunohistochemistry (IHC) in malignant lymphoma is used for selection of patients in clinical trials using brentuximab vedotin, an antibody drug-conjugate targeting the CD30 molecule. For reliable implementation in daily practice and meaningful selection of patients for clinical trials, information on technical variation and interobserver reproducibility of CD30 immunohistochemistry (IHC) staining is required.

Methods and results: We conducted a three-round reproducibility assessment of CD30 scoring for categorised frequency and intensity, including a technical validation, a 'live polling' pre- and post-instruction scoring round and a web-based round including individual scoring with additional IHC information to mimic daily diagnostic practice. Agreement in all three scoring rounds was poor to fair (κ = 0.12–0.35 for CD30-positive tumour cell percentage and κ = 0.16–0.41 for staining intensity), even when allowing for one category of freedom in percentage of tumour cell positivity (κ = 0.30–0.61). The first round with CD30 staining performed in five independent laboratories showed objective differences in staining intensity. In the second round, approximately half the pathologists changed their opinion on CD30 frequency after a discussion on potential pitfalls, highlighting hesitancy in decision-making. Using fictional cut-off points for percentage of tumour cell positivity, agreement was still suboptimal (κ = 0.35–0.60).

Conclusions: Lack of agreement in cases with heterogeneous expression is shown to influence patient eligibility for treatment with brentuximab vedotin, both in clinical practice and within the context of clinical trials, and limits the potential predictive value of the relative frequency of CD30-positive neoplastic cells for clinical response.

Keywords: CD30, immunohistochemistry, interobserver variation, malignant lymphoma

Introduction

Immunohistochemistry (IHC) characterisation is an integral part of daily pathology practice for classifying and subtyping various malignancies, including malignant lymphomas. In recent years, targeted therapies related to specific proteins expressed on tumour cells have prompted the use of IHC for the detection or measurement of these specific molecules as predictive markers for treatment outcome. Examples include human epidermal growth factor receptor 2 (HER2) assessment as a predictive marker for decision-

Address for correspondence: L. Koens, Department of Pathology, Academic Medical Center, Postbus 22660, 1100 DD Amsterdam, the Netherlands. e-mail: l.koens@amc.uva.nl

making in breast cancer treatment with targeted therapy against HER2,[1] programmed cell death ligand 1 (PDL-1) staining on tumour cells and tumour-associated histiocytes in relation to programmed cell death 1 (PD-1) inhibitory treatment in melanoma patients[2] and, increasingly, selection of patients with diffuse large B cell lymphoma (DLBCL) for treatment choices within and outside clinical trials based on IHC algorithms for cell-of-origin classification.[3]

In recent years, CD30 has gained attention as a molecule of interest for targeted therapy of haematological malignancies. CD30 is a type I transmembrane protein with six cysteine-rich pseudo-repeat motifs in its extracellular domain, and contains a cytoplasmic tail with several tumour necrosis factor receptor-binding sequences that are able to activate nuclear factor kappa B (NF-κB) and extracellular signal-regulated kinase signalling pathways.[4] CD30 can be targeted specifically by brentuximab vedotin,[5] a CD30 antibody drug-conjugate, that has shown high efficacy in classical Hodgkin lymphoma (CHL) and anaplastic large cell lymphoma (ALCL), malignant lymphomas with often strong and homogeneous IHC expression of CD30. Other lymphoma classes, such as diffuse large B cell lymphoma (DLBCL) and various T cell lymphoma subtypes [especially extranodal natural killer (NK)/T cell lymphomas and enteropathy-associated T cell lymphomas (EATL)], may express CD30, albeit with heterogeneous staining intensity and percentage of positive tumour cells, and currently the efficacy of treatment with brentuximab vedotin is being explored actively in these lymphoma types.[6,7] However, there is no consensus on CD30 cut-points or the staining pattern that should be observed, and widely variable criteria are used.[8,9]

These developments imply that the role of the pathologist to support selection of patients for treatment will increase further in this field. Building upon the experience with major reproducibility issues and variable cut-off point definitions for predictive IHC markers, both in solid tumours[10,11] and lymphoma,[12] similar challenges may be expected for CD30 testing. Before meaningful implementation of predictive scoring for CD30 in daily practice, this aspect should be evaluated, especially as variations will probably influence eligibility for inclusion in clinical trials and may preclude meaningful correlative studies. Therefore, we performed a three-round formal validation study including aspects of technical reproducibility/interlaboratory variability, interobserver variability and learning effects.

Materials and methods

TISSUE MICROARRAY (TMA)

TMAs were constructed using 20 archival formalin-fixed and paraffin-embedded (FFPE) patient samples originating from one pathology laboratory of various lymphoid malignancies to cover various staining intensities and positive tumour cell frequencies for CD30 and known pitfalls, including 12 cases of DLBCL, three cases of EATL and one case each of mediastinal grey zone lymphoma, adult T cell lymphoma/leukaemia (ATLL), peripheral T cell lymphoma, not otherwise specified (PTCL-NOS), and ALK1-negative ALCL. Two representative 1.0-mm cores were processed using standard procedures.[13] Five-micrometre sections were cut and sent to five pathology laboratories in The Netherlands for staining with CD30 antibodies using local protocols for routine diagnostic procedures.

IHC INTERPRETATION

In all assessments, the percentage of CD30-positive neoplastic cells and the intensity of staining were estimated visually. Positive tumour cells were scored in percentage classes: no expression, > 0–2%, 3–10%, 11–20%, 21–30%, 31–50% and > 50%. Staining intensity was scored as no expression, heterogeneously negative–weak, uniformly weak, heterogeneously weak–strong and uniformly strong.

For round 1 of technical validation and IHC interpretation, each core of the TMA was assessed by the local pathologist of the laboratory that performed the staining procedure ($n = 5$).

Round 2 was performed during a national workshop on CD30 as a therapeutic target in haematological malignancies, in which 25 medical professionals, including (haemato)pathologists, haemato-oncologists and dermatologists, participated in a live polling system for six cases using on-screen photographs of CD30 staining in three cases of DLBCL, two cases of PTCL-NOS and one case of EATL, representative of the spectrum of frequency and intensity of staining. All participants of round 1 were present in scoring round 2. This was followed by a presentation on the pitfalls of CD30 IHC interpretation by one of the authors (L.K.), after which exactly the same scoring procedure was directly repeated. The pitfalls discussed comprised CD30-positivity in reactive cells, technical issues and the interpretation of cases with tumour cells that show the same size as reactive surrounding cells.

Round 3 consisted of 20 cases presented as representative photographs of the haematoxylin and eosin (H&E)-stained slides, CD30 IHC and relevant diagnostic IHC markers. Participants, who had all attended the national workshop, scored the CD30 IHC stain in a series consisting of representative areas of 13 cases of DLBCL, two cases of PTL-NOS, two cases of EATL, one case of ALK1-negative ALCL and one case of extranodal NK/T cell lymphoma, nasal type. All cases were revised beforehand (L.K., D.J.), according to the latest criteria.

STATISTICAL ANALYSIS

Inter-rater agreement was quantified by means of kappa coefficients and percentage of pairs in agreement. Overall kappa coefficients for exact agreement and multiple raters were calculated in STATA version 14[14] for percentage positivity and intensity. Confidence intervals were obtained using a bootstrap procedure. Percentage agreement and two-rater kappa coefficients were calculated in R version 2.3.5[15] for each pair of raters. The average kappa and average percentage agreement were calculated together with their range to show the variability in agreement between different pairs of raters. Kappa coefficients and percentage agreement for percentage positivity allowing for one category of freedom were calculated in R for each pair of raters. The average of the kappa coefficients and their range were calculated. Finally, kappa coefficients and percentage agreement were calculated for positivity using fictional cut-off points of 2 and 10%. We categorised kappas as poor (< 0.40), fair (0.40–0.75) or excellent (> 0.75).

Results

An overview of the results of the three scoring rounds is represented in Table 1.

ROUND 1

IHC for CD30 on a TMA containing 20 lymphoma cases and two staining control tissues was performed in five pathology laboratories according to routine procedures using automated staining protocols [Dako Autostainer platform $n = 2$ (Dako, Glostrup, Denmark), Ventana Medical Systems Benchmark platform $n = 3$ (Ventana Medical Systems, Oro Valley, AZ, USA)] and anti-CD30 antibody clone Ber-H2 [Ventana Ber-H2 (790-4858) $n = 3$, Dako Ber-H2 (IR602) $n = 1$, Dako Ber-H2 (M0751) $n = 1$)]. Slides were scored according to local guidelines. Despite the use of the same antibody clone, the staining results varied dramatically (Figure 1), resulting in pairwise agreement of 46% and a κ of 0.35 for percentage of positive tumour cells and pairwise agreement of 56% and a κ of 0.47 for staining intensity. Overall, there was a minor difference in agreement between the pathologists scoring slides stained in the Dako automated platform (percentage positivity; pairwise agreement 56%/$\kappa = 0.46$ and intensity; pairwise agreement 83%/$\kappa = 0.79$) and those scoring the Ventana platform stained slides (percentage positivity; pairwise agreement 42%/$\kappa = 0.31$ and intensity; pairwise agreement 49%/$\kappa = 0.39$). Different laboratory techniques could not explain the staining and scoring results systematically.

ROUND 2

In round 2, pilot scoring of CD30 was performed as 'real-life validation' using a live polling system with 22 medical professionals. Based on six cases, agreement for all participants was poor both for quantitative results (pairwise agreement 33%/$\kappa = 0.17$) and for assessment of staining intensity (pairwise agreement 53%/$\kappa = 0.36$). Reproducibility was still poor when allowing for one category of freedom in the CD30 tumour cell positivity class [overall pairwise agreement 63%/$\kappa = 0.33$, for (haemato)pathologists pairwise agreement 62%/$\kappa = 0.30$]. The same slides were rescored after a presentation on pitfalls (L.K.), with 17 medical doctors of the first scoring round participating. Sixteen of the 17 participants changed their scores for one to five cases (mean 2.9 cases changed), with one or more categories in either direction or not scoring at all (Figure 2). Fourteen of the 17 participants changed their interpretation of staining intensity in one to six cases, but with a substantially lower mean of 1.8 cases changed. Overall, the changes in interpretation between the two rounds before and after instruction resulted in similar suboptimal agreement scores.

ROUND 3

Round 3 was designed to mimic a true diagnostic situation. Information on classifying lymphoma diagnosis and scanned images of H&E-stained slides and relevant IHC as support for recognition of tumour cells (CD20, CD3) were provided. All cases were scored by 15 participants, including five academic haematopathologists, six pathologists with a special interest in haematopathology and four residents with

Table 1. Overview of the scoring results of the three scoring rounds

		Percentage of tumour cell positivity								Quantification of positivity			
		κ				% agreement				κ		% agreement	
		Exact		1 cat. freedom		Exact		1 cat. freedom		Exact		Exact	
	n	Multi-rater (95% CI)	Range for two-raters	Mean for two-raters	Range for two-raters	Mean for two-raters	Range for two-raters	Mean for two-rater	Range for two-rater	Multi-rater (95% CI)	Range for two-raters	Mean for two-raters	Range for two-raters
Round 1 overall	**6**	**0.34 (0.21–0.47)**	**[0.13, 0.65]**	**0.61**	**[0.25, 0.91]**	**46**	**[27, 71]**	**75**	**[50, 94]**	**0.41 (0.25–0.57)**	**[0.09, 0.79]**	**56**	**[22, 83]**
Round 2.1 overall	**17**	**0.17 (0.096–0.26)**	**[−0.25, 0.79]**	**0.33**	**[−0.60, 1.0]**	**33**	**[0, 83]**	**63**	**[33, 100]**	**0.36 (0.22–0.50)**	**[–0.17, 1.0]**	**53**	**[17, 100]**
Pathologist/ resident pathology	13	0.22 (0.14–0.31)	**[−0.25, 0.79]**	0.30	[−0.20, 1.0]	37	[0, 83]	62	[33,100]	0.32 (0.18–0.45)	[–0.17,1.0]	49	[17,100]
Round 2.2 overall	**17**	**0.14 (0.089–0.197)**	**[−0.39, 1.0]**	**0.34**	**[−0.42, 1.0]**	**29**	**[0, 100]**	**59**	**[0, 100]**	**0.40 (0.28–0.53)**	**[–1.0, 1.0]**	**56**	**[0,100]**
Pathologist	13	0.12 (0.054–0.18)	[−0.39, 0.79]	0.31	[−0.33, 1.0]	26	[0, 83]	57	[0, 100]	0.39 (0.24–0.54)	[–1.0, 1.0]	55	[0, 100]
Round 3 overall	**15**	**0.20 (0.13–0.27)**	**[−0.16, 0.57]**	**0.50**	**[−0.03, 0.92]**	**33**	**[0, 65]**	**71**	**[25, 95]**	**0.16 (0.095–0.23)**	**[–0.15, 0.76]**	**37**	**[5, 85]**
>2% tumour cell positivity	15	0.49 (0.29, 0.69)	[−0.07, 1.0]	x	x	78	[40, 100]	x	x	x	x	x	x
>10% tumor cell positivity	15	0.51 (0.33, 0.70)	[0.11, 0.9]	x	x	76	[45, 95]	x	x	x	x	x	x

CI, confidence interval.
Bold values represent the whole group analyzed per scoring round

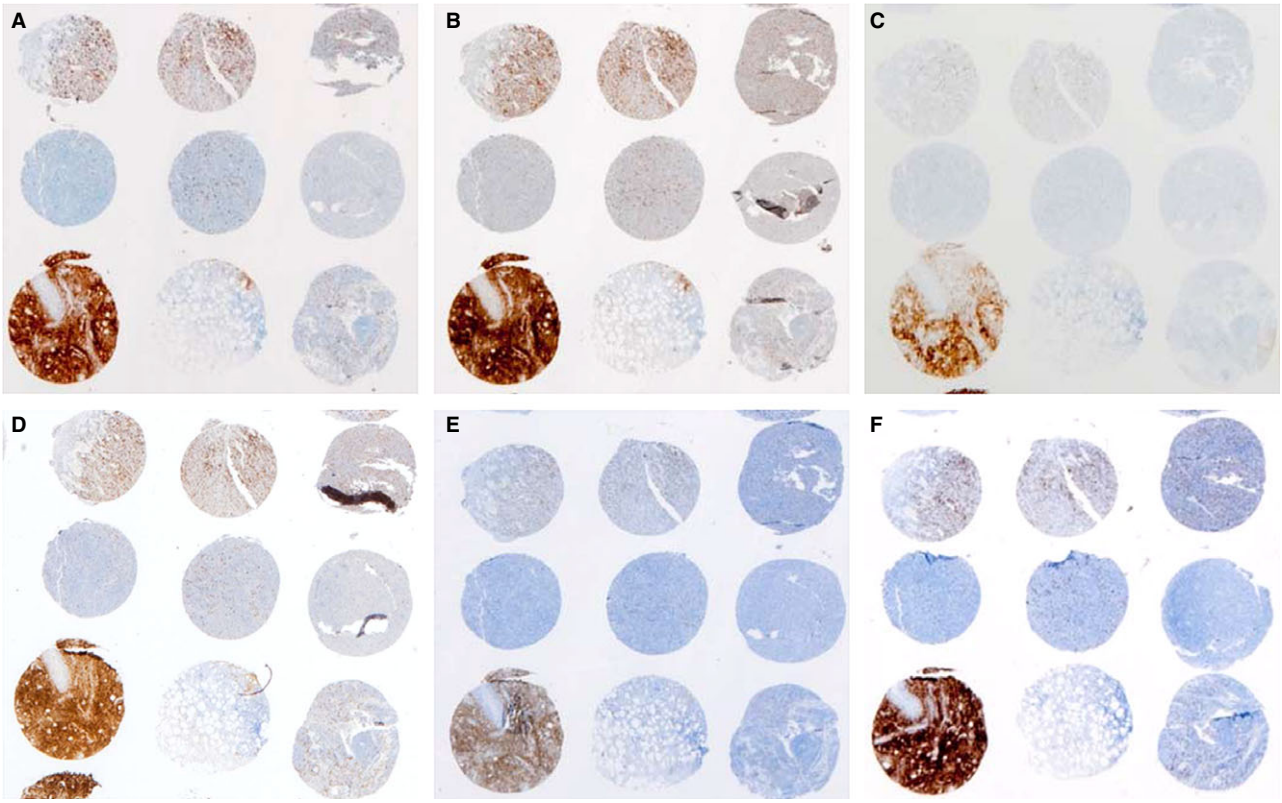

Figure 1. CD30 immunohistochemistry performed on tissue microarrays (TMA) in five different laboratories. An overview of the different CD30 immunohistochemistry slides shows apparent differences in staining intensity in some of the cores. **A**, **B**, Stained by the Ventana Benchmark stainer with a Ventana Ber-H2 antibody; **C**, stained by the same machine, but with a Dako Ber-H2 antibody; **D**, **E**,) stained by the Dako Autostainer, using a Dako Ber-H2 antibody.

basic training in haematopathology. The distribution of percentage classes of CD30-positivity per tumour varied substantially among the individual participants, showing that some pathologists have a systematic tendency for higher scores of CD30-positivity than others (Figure 3). Exact pairwise agreement in CD30-positive tumour cell percentage and staining intensity were 33% ($\kappa = 0.20$) and 74% ($\kappa = 0.36$), respectively, and therefore no substantial improvement from round 2 was reached. In contrast to scoring round 2, allowing for one category of freedom in CD30-positive tumour cell percentage led to an improvement of reproducibility to fair agreement (pairwise agreement 71%/$\kappa = 0.50$). Agreement levels were not dependent upon the level of training or experience in years of practice of the participants.

Using fictional cut-points of 2 and 10% positivity, fair agreement was reached (2% cut-off: pairwise agreement 78%/$\kappa = 0.48$; 10% cut-off: pairwise agreement 76%/$\kappa = 0.52$) (Table 2). A 2% cut-point classified five of 20 cases as positive by all participants, whereas for the 10% cut-off six of 20 cases were scored with complete agreement (three cases CD30-negative and three cases CD30-positive). For implementation of CD30 scoring as a tool for treatment decisions, discordant decisions around the cut off-points are most relevant. Using dichotomised cut-points for (virtual) trial inclusion, opinion on inclusion or not differed from the majority opinion in up to 46% of the pathologists (mean 2.15 participants for the 2% cut-point, mean 2.25 for the 10% cut-point). As an example, in case 15, 11 of 15 pathologists considered the tumour CD30-positive using a 2% cut-point and four of 15 pathologists considered the tumour CD30-positive with a 10% cut-point (Figure 4), emphasising the ambiguity in interpretation, especially in tumours with relatively few CD30-positive tumour cells.

Discussion

Biomarker assays as a selection tool for treatment with targeted compounds should be technically

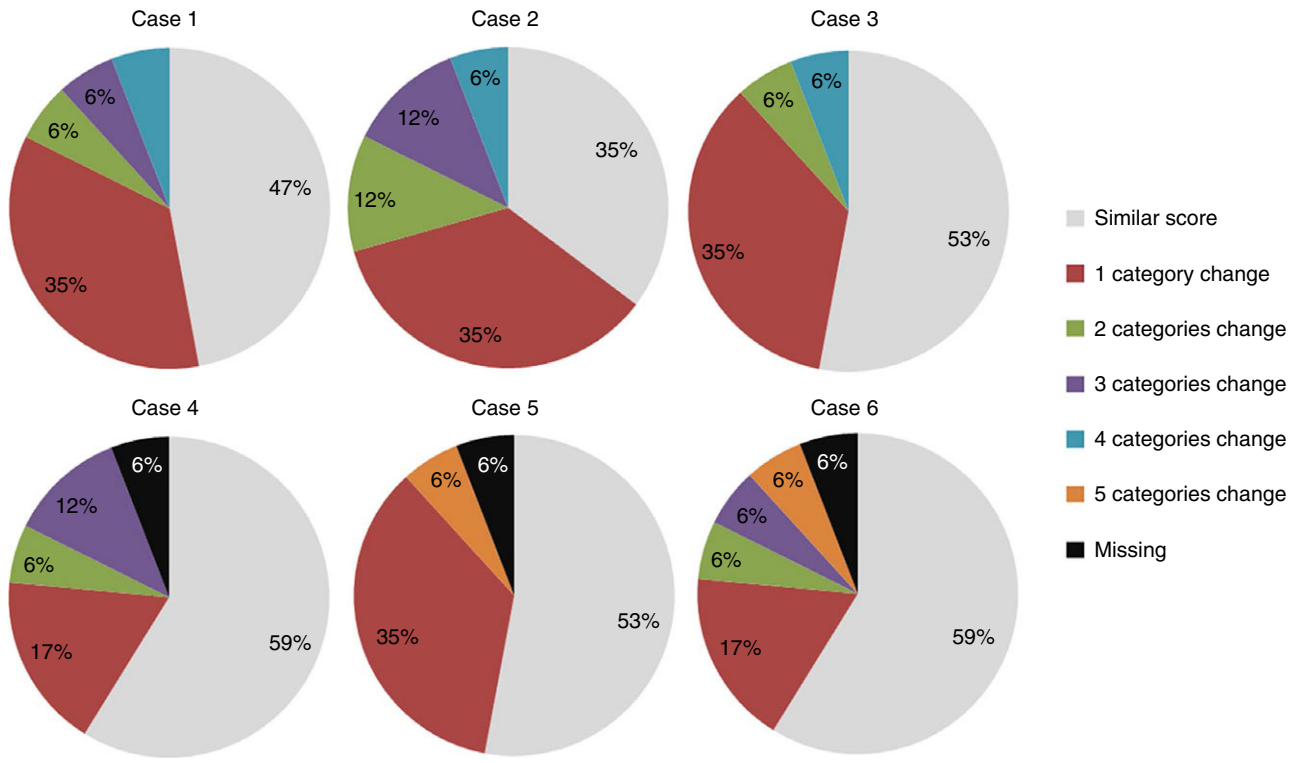

Figure 2. Intra-observer variation in the interpretation of percentage tumour cell positivity. In round 2, a substantial number of the participants changed their opinion on the percentage of tumour cell positivity for the same slide only 15 min after scoring it for the first time, sometimes even changing several scoring categories.

robust and interpretation should be reproducible. In this study, we show that the results for CD30 staining on FFPE biopsy samples of malignant lymphomas are variable between five laboratories, in which this procedure is part of routine lymphoma work-up. Although all results were fully adequate for diagnostic classification purposes, this variation resulted in major differences in quantitative and qualitative assessment of CD30 data. These results are in line with a quality monitoring study by NordiQC, showing that only 179 of 252 (71%) of laboratories tested were able to produce an optimal CD30 staining according to well-described criteria, supporting the notion that staining heterogeneity is a factor that cannot be ignored in the broader pathology community.[13] Technical variation for IHC and its impact on standardisation of biomarker scoring has also been demonstrated for other membranous, cytoplasmic and nuclear markers in lymphoma.[16] As a consequence, we still advise central processing of biopsy samples for treatment selection in the context of clinical trials, including those employing CD30-targeting drugs. However, as tissue fixation and subsequent tissue processing protocols inevitably vary considerably between laboratories, at least some variation will remain inherent to IHC-based assays. It will not be possible to define universally optimised staining procedures as a gold standard for determining CD30-positive tumour cell percentage and intensity.

Variation in CD30-positive tumour cell percentage scoring and intensity assessment cannot only be explained by technical differences between laboratories. Also, when assessing CD30-positive tumour cell percentage and intensity from the same digitalised slides and under the same circumstances, agreement between pathologists is still poor to fair, at best. The difficulty in decision-making was emphasised by the high percentage of participants who showed a high level of intraobserver variability when scoring the same cases twice at the 'real-time validation' effort. Indeed, even experienced haematopathologists in this group were hesitant to provide their scores in the second round after a presentation on pitfalls in interpretation. These results highlight that the same slides can be interpreted in different ways, even by the same pathologist, and interpretation can be influenced by the mention of potential pitfalls. A possible weakness of this 'real-life validation' effort is the

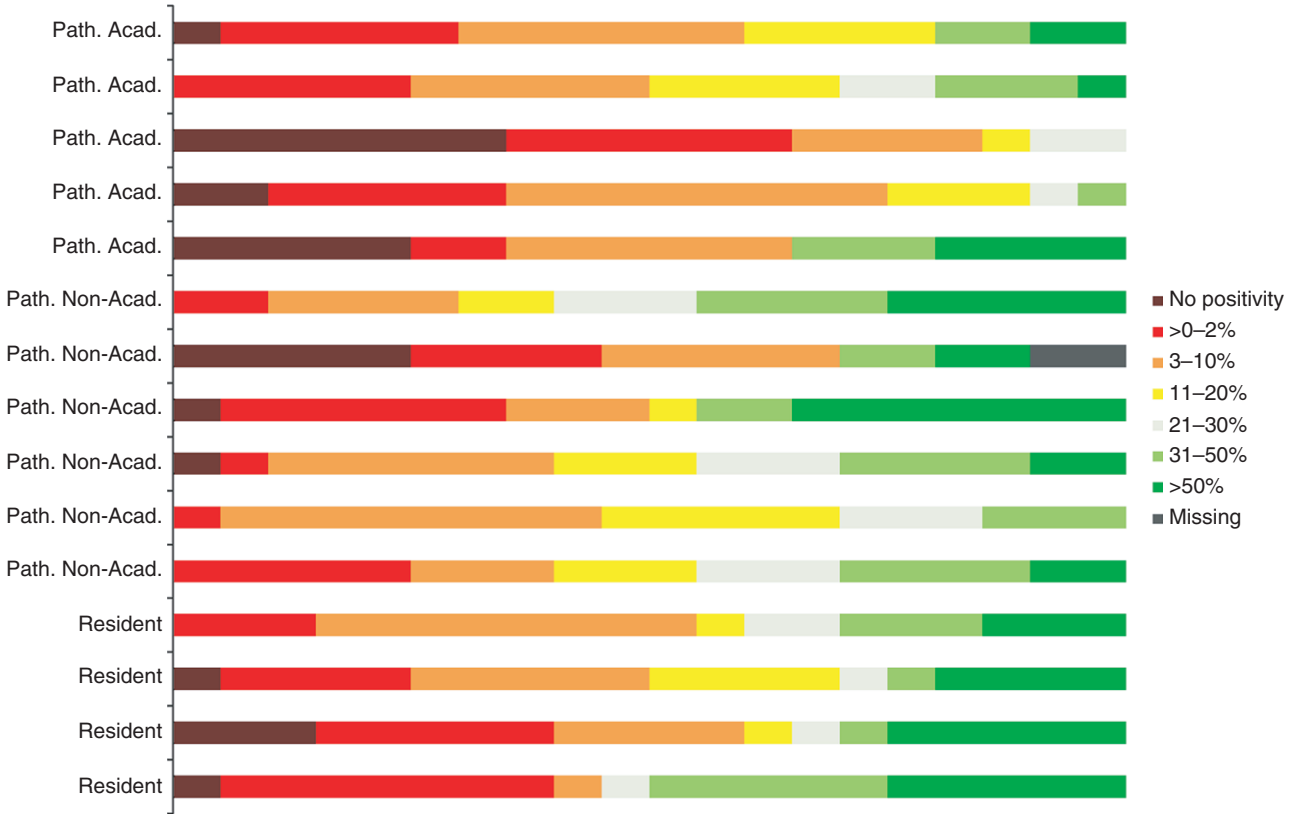

Figure 3. The distribution of scoring CD30 tumour cell positivity percentage. For scoring round 3, the results per individual participants are depicted, emphasising individual variation and the tendency of some participants to easily score higher tumour cell positivity than others.

Table 2. Pairwise agreement and κ using cut-off values for percentage of tumor cell positivity

		> 2% tumour positivity		>10% tumour positivity	
	n	% agreement	κ	% agreement	κ
Round 3 overall	15	78	0.49 (0.29, 0.69)	76	0.51 (0.33, 0.70)
Pathologist	11	80	0.49 (0.27, 0.71)	78	0.54 (0.34, 0.74)
Academic pathologist	5	85	0.57 (0.30, 0.85)	82	0.60 (0.35, 0.85)
Non-academic pathologist	6	76	0.35 (0.14, 0.56)	73	0.45 (0.24, 0.66)
Resident	4	71	0.43 (0.15, 0.70)	75	0.45 (0.12, 0.77)

somewhat artificial setup. In daily practice, IHC stains are never assessed outside their context of clinical information, morphology and a panel of diagnostic immunohistochemical stains to provide information on architectural distribution and cellular properties of tumour cells and reactive cell populations. Therefore, in the third validation round, H&E slides and essential additional images of diagnostic IHC slides were provided to mimic a real-life situation. The agreement did not improve substantially, however. Although the exact agreement in quantifying CD30-positive neoplastic cells was still suboptimal, allowing for one category of freedom in this category improved agreement substantially to fair.

Our study showed that quantifying CD30-positive tumour cells is variable among pathologists. This phenomenon may not pose excessive problems for the majority of patients to be included in clinical trials

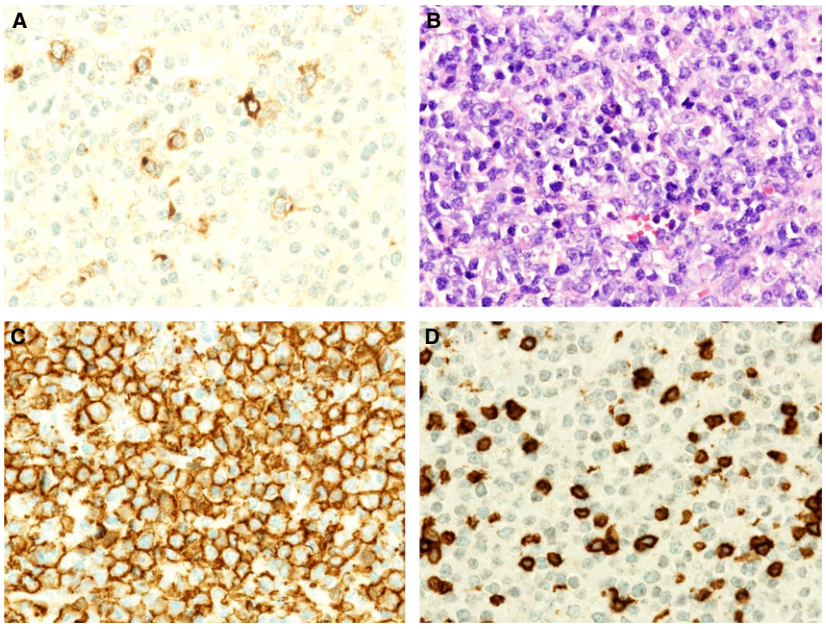

Figure 4. Case example. This case shows the pictures that were evaluated and scored by the participants of a diffuse large B cell lymphoma with CD30 (**A**), haematoxylin and eosin (**B**), CD20 (**C**) and CD3 (**D**), showing considerable variation in assessment, especially using the fictional 2% and 10% cut-off for CD30 positivity.

based on a dichotomised score, as these currently include classes that are uniformly CD30-positive [ALCL; uniform CD30-positivity in 100%, and classical Hodgkin lymphoma; uniform CD30-positivity in 100% and DLBCL, uniform CD30-positivity in 19% in the relapse setting (based on the files of the Amsterdam Comprehensive Cancer Center Database, D. de Jong, personal communication)]. For heterogeneously CD30-positive lymphoma classes that are increasingly considered for targeted treatment, the situation may be more challenging.

One of the alternatives to improve reproducibility of CD30 assessment as a treatment selection tool may be automated image analysis-based scoring. In a Phase II study of brentuximab vedotin in relapsed/refractory DLBCL with variable CD30 expression, all responding patients had quantifiable CD30 by computer-assisted assessment of IHC,[8] albeit that there was no statistical correlation between response and level of CD30 expression. Staining intensity of CD30 was not considered in this study. However, interpretation of IHC stainings, irrespective of conventional 'manual' assessment or computer-assisted scoring, is complicated by the difficult differentiation of CD30 staining in neoplastic cells versus non-malignant CD30-positive cells in the tumour microenvironment, such as various populations of resting CD8-positive T cells, activated T cells, activated reactive B cells and NK cells.[17] In particular, if the cut-off point for CD30-positivity for study eligibility is set at a very low percentage, such as 1 or 2%, reactive CD30-positive cells may easily influence decision-making. In a study in PTCL, CD30 IHC was shown to be correlated highly with mRNA levels using an IHC scoring system incorporating both staining intensity and percentage of positive tumour cells.[18] However, measurement of CD30 mRNA as an alternative assay may be technically more complicated and expensive and, also using this technique, CD30-positive tumour cells cannot be distinguished from CD30-positive surrounding reactive cells. Flow cytometry [fluorescence activated cell sorter (FACS) analysis] has the advantage of a quantitative assay, allows for multiple-marker staining and is often more sensitive than IHC. However, fresh tissue suspensions, necessary for this technique, are not always available and the cell membrane of the large tumour cells of CD30-positive T and B cell lymphomas is often vulnerable and easily shed when preparing cell suspensions for FACS, precluding use in daily practice.[19] Another way to evaluate CD30 is the detection of soluble CD30 in peripheral blood. Soluble CD30 is the extracellular domain of CD30 that is released into the circulation after proteolytic cleavage near the cell membrane, and can be detected by enzyme-linked immunosorbent assay (ELISA).[20] Soluble CD30 levels have been shown to be correlated

with disease burden in ALCL,[21] clinical features and prognosis in CHL,[22] but the levels of soluble CD30 are not correlated with clinical response to brentuximab vedotin in relapsed/refractory DLBCL.[8] These alternative methods for CD30 quantification therefore all seem to have more disadvantages than benefits, and conventional visual scoring of CD30 IHC by pathologists thus remains an important method to be optimised.

The role of staining intensity of CD30 in the clinical response to treatment with brentuximab vedotin is unclear. The only study correlating CD30 expression with this response did not consider staining intensity.[8] The study showing high correlation between CD30 IHC and CD30 mRNA levels[18] considered both staining intensity and percentage of positive tumour cells, indicating that staining intensity might be extremely relevant in assessing this marker. This study was, however, restricted to peripheral T cell lymphomas, and there is no evidence that this type of CD30 IHC scoring or mRNA expression are correlated with clinical response to brentuximab vedotin.

In summary, reproducibility of the IHC CD30 stain is suboptimal, in part by variation in staining methods and patterns between different pathology laboratories, but due also to interobserver variation between pathologists. These differences could potentially influence patient eligibility for clinical trials with antibody-drug conjugate brentuximab vedotin, and also hamper the correlation of the amount of CD30-positive neoplastic cells to the degree of clinical response to this treatment.

Acknowledgements

The authors thank Takeda Netherlands for financial support for the 'National Workshop on CD30 as a Therapeutical Target in Hematological Malignancies', during which round 2 of this validation study was conducted. We thank all participants for their time and efforts. We thank Sander Veltkamp for stimulating advice.

Conflicts of interest

The authors state no conflicts of interest.

References

1. Wolff AC, Hammond ME, Hicks DG *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* 2013; **31**; 3997–4013.
2. Daud AI, Wolchok JD, Robert C *et al.* Programmed death-ligand 1 expression and response to the anti-programmed death 1 antibody pembrolizumab in melanoma. *J. Clin. Oncol.* 2016; **34**; 4102–4109.
3. Scott DW. Cell-of-origin in diffuse large B-cell lymphoma: are the assays ready for the clinic? *Am. Soc. Clin. Oncol. Educ. Book* 2015; **35**; e458–e466.
4. Younes A. CD30-targeted antibody therapy. *Curr. Opin. Oncol.* 2011; **23**; 587–593.
5. Francisco JA, Cerveny CG, Meyer DL *et al.* cAC10-vcMMAE, an anti-CD30-monomethyl auristatin E conjugate with potent and selective antitumor activity. *Blood* 2003; **102**; 1458–1465.
6. Slack GW, Steidl C, Sehn LH, Gascoyne RD. CD30 expression in *de novo* diffuse large B-cell lymphoma: a population-based study from British Columbia. *Br. J. Haematol.* 2014; **167**; 608–617.
7. Sabattini E, Pizzi M, Tabanelli V *et al.* CD30 expression in peripheral T-cell lymphomas. *Haematologica* 2013; **98**; e81–e82.
8. Jacobsen ED, Sharman JP, Oki Y *et al.* Brentuximab vedotin demonstrates objective responses in a phase 2 study of relapsed/refractory DLBCL with variable CD30 expression. *Blood* 2015; **125**; 1394–1402.
9. Weisenburger DD, Savage KJ, Harris NL *et al.* Peripheral T-cell lymphoma, not otherwise specified: a report of 340 cases from the International Peripheral T-cell Lymphoma Project. *Blood* 2011; **117**; 3402–3408.
10. Perez EA, Suman VJ, Davidson NE *et al.* HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial. *J. Clin. Oncol.* 2006; **24**; 3032–3038.
11. Ruschoff J, Kerr KM, Grote HJ *et al.* Reproducibility of immunohistochemical scoring for epidermal growth factor receptor expression in non-small cell lung cancer: round robin test. *Arch. Pathol. Lab. Med.* 2013; **137**; 1255–1261.
12. de Jong D, Rosenwald A, Chhanabhai M *et al.* Immunohistochemical prognostic markers in diffuse large B-cell lymphoma: validation of tissue microarray as a prerequisite for broad clinical applications – a study from the Lunenburg Lymphoma Biomarker Consortium. *J. Clin. Oncol.* 2007; **25**; 805–812.
13. Nordic Immunohistochemical Quality Control. Recommended CD30 protocols. 2015. updated 1 October 2017. Available at: http://www.nordiqc.org/epitope.php?id = 5.
14. StataCorp. *Stata statistical software*, release 14 edn. College Station, TX: StataCorp LP, 2015.
15. R Development Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2010.
16. Sander B, de Jong D, Rosenwald A *et al.* The reliability of immunohistochemical analysis of the tumor microenvironment in follicular lymphoma: a validation study from the Lunenburg Lymphoma Biomarker Consortium. *Haematologica* 2014; **99**; 715–725.
17. Horie R, Watanabe T. CD30: expression and function in health and disease. *Semin. Immunol.* 1998; **10**; 457–470.
18. Bossard C, Dobay MP, Parrens M *et al.* Immunohistochemistry as a valuable tool to assess CD30 expression in peripheral T-cell lymphomas: high correlation with mRNA levels. *Blood* 2014; **124**; 2983–2986.

19. de Tute RM. Flow cytometry and its use in the diagnosis and management of mature lymphoid malignancies. *Histopathology* 2011; **58**; 90–105.
20. Josimovic-Alasevic O, Durkop H, Schwarting R, Backe E, Stein H, Diamantstein T. Ki-1 (CD30) antigen is released by Ki-1-positive tumor cells *in vitro* and *in vivo*. I. Partial characterization of soluble Ki-1 antigen and detection of the antigen in cell culture supernatants and in serum by an enzyme-linked immunosorbent assay. *Eur. J. Immunol.* 1989; **19**; 157–162.
21. Nadali G, Vinante F, Stein H *et al.* Serum levels of the soluble form of CD30 molecule as a tumor marker in CD30 + anaplastic large-cell lymphoma. *J. Clin. Oncol.* 1995; **13**; 1355–1360.
22. Nadali G, Vinante F, Ambrosetti A *et al.* Serum levels of soluble CD30 are elevated in the majority of untreated patients with Hodgkin's disease and correlate with clinical features and prognosis. *J. Clin. Oncol.* 1994; **12**; 793–797.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Round 3: the photographs of CD30 immunohistochemistry and additional slides and diagnosis information